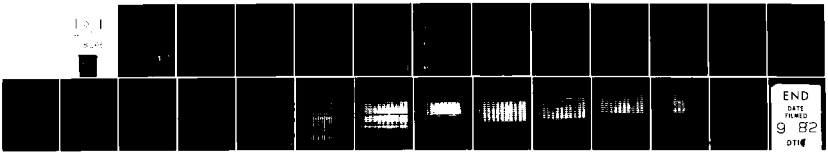
AD-A118 265 DUKE UNIV DURHAM NC CENTER FOR APPLIED BUSINESS RESEARCH F/G 5/9
CONFIDENCE INTERVALS AND VALIDATION OF A FORECASTER OF QUALITY --ETC(U)
JUL 82 R C MOREY, J M MCCANN N00014-80-C-0200

UNCLASSIFIED NL



A Report Prepared for Naval Recruiting Command
and the Office of Naval Research Under Contract N000-14-80-C-0200

Report ONR-200-6

CONFIDENCE INTERVALS AND VALIDATION OF A FORECASTER OF QUALITY NAVY ENLISTMENTS

July, 1982

Principal Investigator: Richard C. Morey, Ph.D.
Associate: John M. McCann, Ph.D.

CENTER FOR APPLIED BUSINESS RESEARCH
FUQUA SCHOOL OF BUSINESS
DUKE UNIVERSITY
DURHAM, NORTH CAROLINA
(919) 684-2012

TABLE OF CONTENTS

1.0 INTRODUCTION AND SUMMARY

1.1 Background

A substantial amount of effort has been expended in the last few years ~~(e.g., see [1],~~ [2], [4]) to attempt to be able to improve the forecasting of the number of quality recruits that will enter the various services over some given period; the "quality" label refers to those supply limited recruits with High School Diplomas and/or those scoring in the Upper Mental categories on the AFEES exams. The explanatory variables used in the forecasting models are the levels of key resources such as recruiters and advertising levels of different types, and key demographics such as the unemployment rate, the number of male high school seniors, etc. As numerous and diverse as these efforts have been, very few have been subjected to rigorous types of validation and none, to the knowledge of the authors, has yielded rigorous confidence intervals. This deficiency was one of the major criticisms of the discussants at the ONR Personnel Supply Models Workshop in late January, 1981 (see [6]). The need for statistical confidence intervals is, of course, to put into proper perspective the single point estimates generated, and to quantify the uncertainty or risk remaining. Decision makers are then in a position to apply their own risk preferences and to factor in the non-quantifiable considerations.

This report discusses the result of two separate validation efforts and the development of statistical confidence intervals for a predictor of quality enlistment contracts that is currently being used by the Naval Recruiting Command. The forecaster used for aiding in budget generation, resource allocation and the spreading of the quotas over the districts. The predictor is for the number of male, non-prior service, active duty, High School Graduate (including those with GED's) contract enlistments obtained in a given district, Region or the nation over a given period

of time. It is a non-linear function of thirteen explanatory variables plus a number of monthly indicator variables. The model was built using monthly, district data from January 1976 to December 1978 and contains 1,548 cells. The key explanatory variables are: the local unemployment rate by district by month; the number of production recruiters by district by month; the real dollars of advertising in local advertising; the total dollars spent in the Navy's so called General Enlisted Program-General (GEP-G) budget on television, radio and billboards; the expenditures in the GEP-G budget on printed materials (i.e., direct mail, magazines, newspapers); the total expenditures in the Navy's GEP-Minority Program, the total advertising dollars spent in the Joint Armed Forces Program (JADOR); the ratio of the first year military pay to the average civilian pay for non-agricultural work and non-supervisory personnel; the urban-rural character of the district; the number of male High School seniors in the district; and the percent black in the male, 17-21 year old population of the district.

A two equation system was used where NOIC leads were first predicted as a function of the advertising and demographic variables; the second equation was on HSG male, non-prior contracts where NOIC leads was an explanatory factor. A log-log model was used to capture the diminishing return nature of recruiting resources. A Koyck autoregressive term (e.g., see [3]) was used to account for the lagged effects of advertising, and the so-called Park's method of regression (e.g., see [3]) was used to handle the strong autocorrelation and heteroscedacity associated with pooling the time series and cross-sectional data. More details are available in the authors' ONR report of July 1980, entitled, "The Impacts of Various Types of Advertising Media, Demographics, and Recruiters on Quality Enlistments," (see [4]).

1.2 Results of Validation Efforts

The final model was subjected to two "validation" tests for the independent periods of January 1979 - September 1979, and FY80 in a retrospective mode. The levels of resources and demographics were known with certainty since the validation work was performed in 1982. The final two system model, with its thirteen explanatory variables and monthly indicator variables, was first used to forecast NOIC leads; then HSG contracts were forecasted using the predicted NOIC leads as an explanatory variable. The results were very encouraging:

i) for the independent 9 month period January 1979 - September 1979 (independent in the sense that the model was built using data from other time periods), the model underpredicted the national 9 month totals but by only 3.7%.

ii) For the complete fiscal year FY80, i.e., October 1979 - September 1980, it underpredicted again, but by only 2.5%.

The model also functioned reasonably well at the Regional and monthly levels. The disaggregated results for the last 9 months of FY79 are included.

TABLE 1

DEGREE OF AGREEMENT BETWEEN PREDICTED
AND ACTUALS FOR THE INDEPENDENT
TIME PERIOD OF 1/79-9/79

High School Graduate Contracts

	Actual	Predicted	(Error Rate*)
NATIONAL ENTIRE 9 MONTH PERIOD	45,137	43,459	(-3.7%)
Area, Entire 9 Month Period			
Area 100	9,667	9,102	
Area 300	8,582	8446	
Area 400	8,409	8,321	
Area 500	4,986	5,005	
Area 700	5,592	5,531	
Area 800	7,901	7,004	
MEAN ABSOLUTE ERROR RATES OVER ALL AREAS	3.7%		
NATIONAL MONTHLY RESULTS			
January 1979	5,316	5,150	
February 1979	4,639	5,463	
March 1979	4,931	5,519	
April 1979	4,030	4,113	
May 1979	4.007	3,772	
June 1979	5,361	4,802	
July 1979	5,416	4,951	
August 1979	6,313	4,975	
September, 1979	5,124	4,715	
MEAN ABSOLUTE ERROR RATE OVER ALL MONTHS	9.87%		

* All error rates are $\frac{\text{"Predicted"} - \text{"Actuals"}}{\text{"Actuals"}}$

1.3 <u>Calculation of a Statistical Confidence Interval for the National, Annual Level of HSG Graduate) Contracts</u> (male, non-prior service, active duty)

One of the reasons that other researchers have not generated confidence intervals is the complexity of the issues involved:

i) the predictors typically include non-linearities to capture the diminishing return nature of recruiting resources;

ii) the predictors, in order to capture the "good will" effect of advertising must necessarily include lagged variables, and hence colinearities are introduced;

iii) the predictor must include monthly or quarterly seasonal variables to reflect the seasonal nature of recruiting.

iv) the predictors typically exhibit error terms which are highly correlated across districts, have unequal variances (heteroscedasity) and are autocorrelated.

The basic approach described subsequently in Section 3 relies on detailed information provided by the so called Park's regression package, available from the SAS software. Unlike Ordinary Least Squares (OLS) regression packages, it is geared to handle, quantify and incorporate the above effects. It provides a great wealth of information that can be used to generate rigorous confidence intervals which deal with the considerations in (i) - (iv).

When this was accomplished, the following results were yielded: assuming the demographics and resources to be utilized are forecasted properly, then at the national level, <u>one can be 90% confident the actual level of male, non-prior service, HSG contracts will fall within ±8% of the single point predicted level</u>; <u>if a confidence factor of 80% is used, the interval is ±6.3%.</u>

The confidence level for each of the Navy's six Recruiting Areas follows below. It is noted that they are less precise since there is a considerable amount

of smoothing or averaging obtained when working at the national level, in contrast to the Regional level. When one appreciates that the regression model is being forced to fit all districts and regions (with no dummy or indicator variables of any kind being included), the results are reasonable. Subsequent research is being geared to developing separate predictive equations with separate elasticity estimates for each Region. This should substantially reduce the uncertainties at the Regional levels.

Confidence Limits by Area

	90%	80%
Area 100	±16.6%	±12.9%
Area 300	±14.8%	±11.5%
Area 400	±22.6%	±17.6%
Area 500	±19.3%	±15.0%
Area 700	±13.1%	±10.2%
Area 800	±20.8%	±16.2%
Nation	± 8.0%	± 6.3%

2.0 THE PARK'S REGRESSION MODEL AND ITS OUTPUTS FOR THE HSG CONTRACT PREDICTOR

The HSG contract equation referred to earlier was built using pooled data from 43 districts and 36 months; autocorrelation of the error terms and unequal variances of the error or disturbance terms were observed. In such situations the assumptions underlying the traditional Ordinary Least Squares (OLS) regression techniques are not satisfied and hence OLS is not a viable describes the Park's approach for handling the above problems which is available on the SAS Software. A simple version of the Park's approach is shown below:

$$Y_{it} = \alpha_0 + \alpha_1 X_{it} + \varepsilon_{it} \qquad (i = 1, 2, \ldots, 43;\ t = 1, 2, \ldots, 12)$$

$$\varepsilon_{it} = \rho_i \varepsilon_{i,t-1} + U_{it} \qquad (i = 1, 2, \ldots, 43;\ t = 1, 2, \ldots, 12)$$

$$\mathrm{Var}(\varepsilon_{it}) = \sigma_i^2 \qquad (i = 1, 2, \ldots, 43)$$

$$\text{Covariance }(\varepsilon_{it}, \varepsilon_{j,t+v}) = \begin{cases} \sigma_{ij} \text{ if } v = 0 \\ 0 \text{ otherwise} \end{cases}$$

The other assumptions are that the U_{it} are normally distributed with mean 0 and variance ϕ_{ii}, the covariance of the (U_{it}, U_{jt}) matrix is ϕ_{ij}. Finally, the initial error terms $\varepsilon_{i,o}$ are normally distributed with mean 0 and variance $\phi_{ii}/1 - \rho_i^2$, and have a covariance matrix $E(\varepsilon_{io} \cdot \varepsilon_{jo})$, given by $\phi_{ij}/1 - \rho_i \rho_j$.

Hence in summary the disturbances are allowed to be first-order autocorrelated, i.e., ε_{it} is correlated with $\varepsilon_{i,t-1}$, with a unique (for each district) autocorrelation coefficient ρ_i. Further the disturbances are contemporaneously correlated across the districts (i.e., ε_{it} and ε_{jt} are correlated). Also note the variance of the error term can be different for each district, i.e., it is not necessarily the case that $\sigma_i^2 = \sigma^2$.

Consider first the traditional outputs for the HSG contract predictor described earlier. Since the regression model is a log-log model, the beta values in Table 1 can be interpreted as the short term elasticities.

TABLE 2

Explanatory Variable	Estimated Beta Value	Estimated Standard Errors	t Value
1) Ratio of Military Pay to Civilian Pay	.1583	.014	11.149
2) Number of Male High School Seniors	.2314	.019	12.079
3) NOIC Leads from 2 Months Earlier	.009	.002	4.3998
4) Military Propensity (proxy for proximity of military bases and tradition of military in area; based on responses from a questionnaire)	.6312	.022	28.567
5) Percent Blacks of the 17-21 Year Old, Male Population in the District	-.0007	.004	-.16
6) Urban-Rural Character of the District, (percent of male 17-21 year old population of the district residing in a SMSA)	.183	.0096	19.25
7) Local Advertising Expenditures (deflated so dollars represent constant purchasing power)	.0427	.0058	7.34
8) Number of Production Recruiters in District	.6855	.0145	47.249
9) Local General Unemployment Rate	.1706	.0107	15.925
10) Koyck Autoregressive Term	.0569	.003	14.718

In addition to the above, there were eleven monthly dummies, two year dummies, a GI Bill dummy for the month of December, 1976 (when the GI Bill terminated) and 2 dummies representing the changes in the advertising policy of the Recruiting Command. We note that the beta's obtained from this Park's are different than those obtained from the OLS model which assumes that ρ_i are 0 and the σ_{ii}^2 are all the same. The R^2 of the model is .837.

Next consider the special types of information provided by the Park's Model. Consider first the estimates of the autocorrelation coefficients ρ_i ($i = 1, 2, \ldots, 43$). They are given in Appendix 1 and range from -.3133 (for the Atlanta district) to .598 (for the Little Rock district). We further observe that eleven of the forty-three ρ_i's are negative, and that thirty-two of them have an absolute value larger than .1. Hence it is clear that the error terms or residuals are strongly correlated over time, as might well be expected.

Next consider variance-coveriance matrix of the beta matrix, the beta's being the regression estimates. Since there are twenty-six explanatory variables plus the intercept, this is a 27 x 27 matrix and includes the variances of the estimates (i.e., the square of the standard errors of the estimates) as well as the correlations between the parameters being estimated. As an example, the estimate of the unemployment elasticity (a random variable) has a mean of .1706, a variance of .0001147, and a covariance with the elasticity estimate for the percent black of -.00000829 (i.e., a correlation of -.16397) a covariance with the elasticity of the urban-rural factor of -.00000809 (i.e., a correlation of -.078) and a covariance with the number of production recruiters of .00000028 (i.e., a correlation of .018236). These types of information are needed in generating the confidence intervals sought for. It is also extremly useful in resource allocation decisions where one wishes to develop a confidence interval for the ratio of two elasticities (see [5]) for an application of these ideas to the development of a confidence interval for the optimal ratio of print to non-print advertising as to maximize NOIC leads). The detailed 27 x 27 variance-covariance matrix is shown in Appendix 2.

Finally, consider the ϕ_{ij} ($i = 1, 2, \ldots, 43$; $j = 1, 2, \ldots, 43$) where ϕ_{ij} is the variance-covariance matrix of the (U_{it}, U_{jt}) where

$$\epsilon_{it} = \rho_i \epsilon_{i,t-1} + U_{it}. \tag{1}$$

This matrix captures the variances and contemporaneous correlation between districts of the error terms and is again needed in the confidence interval calculations. As an illustration, the $\emptyset_{ii}$'s range from .02 for the Boston district to .38 for the Louisville district, with most of them in the range of .06 to .08. The entire 43 x 43 matrix is included in Appendix 3. Armed with the information from above, we are in a position to calculate the confidence intervals.

3.0 TECHNIQUE FOR CALCULATING A CONDITIONAL CONFIDENCE INTERVAL FOR PREDICTOR OF MALE HSG CONTRACTS

3.1 Sources of Uncertainty

There are always two sources of uncertainty in using a predictor: the first is that the regression parameter estimates (i.e., the beta's) are random variables which are not known with certainty, that is, the point estimates of the beta's could be in error. The second source of uncertainty is that the values of the independent or explanatory variables may not be known perfectly, i.e., the unemployment rate for next year by district, the number of recruiters, the number of male High School seniors, etc. In this effort, we will assume the values of the explanatory variables are known with certainty and derive a confidence interval which deals with the first type of uncertainty only; this is known as a conditional confidence interval, conditional on the values of the X's being known.

3.2 Overview of Approach

The basic approach is one of Monte Carlo simulation where the realizations will be drawn from random variables which are dependent on the three types of information provided by the Park's Model. Recall that the model is:

$$Y_{it} = \beta_o + \sum_{k=1}^{26} \beta_k X_{kit} + \epsilon_{it} \quad (i = 1, 2, \ldots, 43;\ t = 1, 2, \ldots, 12) \tag{2}$$

where Y_{it} is the log of the number of male, non-prior service HSG contracts and where X_{kit} is the log of the various explanatory factors; the β_k are then the true but unknown elasticities, i.e., β_k represents the percent change in HSG contracts of a 1% change in the factor X_k.

Consider the simulation for the district of Albany, N.Y., i.e., i = 1 where the key output of the simulation will be realizations of Y_{it} and finally a confidence interval for the first month of the fiscal year, i.e., October. The X values will be the values of the independent variables for FY80. Then for Albany, October, 1979, we know all of the X's (i.e., the number of High School seniors, recruiters, unemployment rate, the monthly indicator variables, the number of NOIC leads for two months ago, and the number of High School Graduate contracts for the previous month, i.e., for September, 1979). In order to generate a realization for $Y_{\text{Albany, October 1979}}$, we will need a realization of the β vector (27 of them) and a realization for $\epsilon_{\text{Albany, October, 1979}}$. Given these realizations and the X's for Albany, October 1979 then from (2) we will have a realization for $Y_{\text{Albany, October, 1979}}$. If one does this say 100 times one has 100 realizations for $Y_{\text{Albany, October, 1979}}$. One can then develop say a $(1 - \alpha)$th of confidence interval for the predictor for Albany, October, 1979 by computing the sample standard deviation, call it $\hat{\sigma}$ and using the forecasted level, call it $\hat{Y} + \hat{\sigma}\phi^{-1}(1 - \alpha/2)$, where $\phi^{-1}(1 - \alpha/2)$ is the $(1 - \alpha/2)$th percentile from the normal distribution.

The general approach is to repeat this procedure for each of the 43 districts and for each of the twelve months, thereby coming up with 100 realizations for the number of High School Graduate contracts for the nation for FY80. By again computing the sample standard deviation for the annual national totals, a confidence interval for the number of national, yearly male, non-prior service High School Graduate contracts is obtained.

3.3 The Detailed Steps

Returning to the details, consider the simulation for Albany for October, 1979.

The steps are as follows:

i) First draw 100 realizations of the β vector (each realization having 27 components). This is done by drawing from a multivariate normal distribution (27 variates) with means given by the point estimates of the β's (i.e., the estimated β values of Table 1) and with a variance-covariance matrix equal to that shown in Appendix 2 (i.e., the 27 x 27 matrix of the Beta values). A standard Monte Carlo technique will yield random samples from a given multivariate normal distribution; this yields the 100 realizations for the vector of β values. The only remaining task is to generate a random draw of $\varepsilon_{\text{Albany, October, 1979}}$ (the error terms). Recall from (1) that:

$$\varepsilon_{\text{Albany, October, 1979}} = \rho_{\text{Albany}}\, \varepsilon_{\text{Albany, September, 1979}} + U_{\text{Albany, October, 1979}}$$

ii) Now ρ_{Albany} is the first entry of the table of autocorrelation coefficients, shown in Appendix 1.

iii) Consider the draw from $\varepsilon_{\text{Albany, September 1979}}$. In the Park's discussion of Section 2, it was pointed out that $\varepsilon_{i,o}$ is normally distributed with mean 0 and variance $\phi_{ii}/1 - \rho_i^2$. Hence $\varepsilon_{\text{Albany, September, 1979}}$ is normally distributed with mean 0 and variance $\phi_{1,1}/1 - \rho_1^2$ where $\phi_{1,1}$ is the first entry in the 43 x 43 matrix of Appendix 3. Hence by making 100 random draws from a normal with mean 0 and variance $\phi_{1,1}/1 - \rho_1^2$, we have 100 realizations of $\varepsilon_{\text{Albany, September, 79}}$

iv) Consider the 100 realizations needed of $U_{\text{Albany, October, 1979}}$, i.e., of $U_{1,1}$. Recall that U_{it} ($i = 1, 2, \ldots, 43$) is assumed to be a 43 variate multivariate normal with means 0 and a 43 x 43 variance-covariance given by ϕ_{ij} (the entries in Appendix 3) which is invariant for all t. Hence by making a 100 draws from a 43 variate normal with the above means and variance-covariance matrix, we have a 100 realization for $U_{1,1}$ which reflects the pairwise correlations (i.e., the contemporaseous correlations).

v) Combining the draw of $U_{1,1}$ with the draw for $\epsilon_{1,0}$ and ρ_1, we have a draw (i.e., a random realization) for $\epsilon_{1,1}$. Combining this with a draw of the β vector and the known value of the X vector, we have a realization for $Y_{1,1}$.

This procedure is repeated for every district and for every month (the same 100 β values can be reused as the Beta's are assumed to hold for every district and every month). For the downstream months, where the actual level of NOIC leads obtained 2 months ago is not known, the forecasted levels from the NOIC regression model is used. (Recall that a NOIC regression model was also developed as part of the 2 equation system.) In addition, whenever the model calls for the number of High School Graduate contracts obtained in the previous month, the forecast of the value obtained for the previous month is used. In this leap-frog manner, all of the simulations can be carried out. The end result of this exercise are 100 realizations for the number of male High School Graduate contracts obtained nationwide in FY80.

4.0 RESULTS

4.1 Results for the Nation

The sample standard deviations from the 100 random realizations, is 3,038. The national prediction for FY80, based on the point estimates for each of the monthly-district pairs, was 62,306 (or 2.5% less than the actual 63,929). Hence the 90% conficence interval is given by 62,306 $\pm$ 1.645 (3,038) or an interval of about $\pm$8%. The 80% interval ($\pm$ 1.282 standard deviations) is $\pm$6.3% whereas the 95% interval (i.e., $\pm$1.96 standard deviations) is about 9.5%.

4.2 Results by Area

The procedure was performed separately for each of the Recruiting Command's six Areas to help discern which Regions were best fitted by the single model and where the largest uncertainties still remained. The resulting six sample standard deviations were:

Area	Forecasted Level of HSG Contracts For FY80	Actual Level of HSG Contracts For FY80	80% Confidence Interval	90% Confidence Interval	95% Confidence Interval
100	12,362	12,799	10,765-13,957	10,314-14,409	9,922-14,801
300	12,385	11,053	10,956-13,814	10,552-14,219	10,200-14,470
400	11,528	13,508	9,497-13,560	8,922-14,135	8,422-14,635
500	8,393	8,499	7,130-9,665	6,771-10,024	6,460-10,334
700	8,343	7,333	7,495-9,192	7,254-9,432	7,046-9,641
800	9,295	10,737	7,784-10,805	7,356-11,233	6,985-11,604

We observe that for every region, the actual level of contracts fell in the 90% and 95% confidence interval. Also, for all but one region, i.e., Area 700, the actual level of HSG contracts fell within the narrowest interval, i.e., the one of 80%. This also helps to instill some credibility in the use of the above confidence intervals.

5.0 REFERENCES

1. Fernandez, Richard, "Enlisted Supply in the 1980's," Rand Corporation Report, WD-515-MRAL, February, 1980.

2. Goldberg, Lawrence, "Recruiters, Advertising and Navy Enlistments," Center for Naval Analysis Report, March, 1980.

3. Kmenta, Jan, Elements of Econometrics, MacMillan Company, New York, 1971.

4. Morey, Richard C. and McCann, John, "The Impacts of Various Types of Advertising Media, Demographics and Recruiters on Quality Enlistments," Office of Naval Research Technical Report, December, 1980.

5. Morey, Richard C. and McCann, John, "Optimal Marketing Mix and Uncertainty: An Application to Lead Generation," Duke University Report, July, 1982.

6. Sinaiko, Wallace H., Miller, J.J. and Cirie, Jack, "Personnel Supply Models Workshop Report, Office of Navy Research Report, 1981.

6.0 APPENDICES

APPENDIX 1

ESTIMATED VALUES OF RHO(I)
(FIRST ORDER AUTOREGRESSIVE PARAMETER ESTIMATES)

I	= 1	-0.050002
I	= 2	-0.123406
I	= 3	[illegible]
I	= 4	[illegible]
I	= 5	0.14087
I	= 6	0.27967
I	= 7	[illegible]
I	= 8	[illegible]
I	= 9	0.24046
I	= 10	0.37415
I	= 11	[illegible]
I	= 12	[illegible]
I	= 13	0.42511
I	= 14	0.10143
I	= 15	[illegible]
I	= 16	[illegible]
I	= 17	0.46318
I	= 18	0.37881
I	= 19	[illegible]
I	= 20	[illegible]
I	= 21	0.38264
I	= 22	0.58637
I	= 23	[illegible]
I	= 24	[illegible]
I	= 25	0.45418
I	= 26	-0.081443
I	= 27	-0.046209
I	= 28	[illegible]
I	= 29	-0.018254
I	= 30	0.33506
I	= 31	[illegible]
I	= 32	[illegible]
I	= 33	0.44679
I	= 34	0.38249
I	= 35	[illegible]
I	= 36	[illegible]
I	= 37	0.35052
I	= 38	-0.081974
I	= 39	[illegible]
I	= 40	[illegible]
I	= 41	[illegible]
I	= 42	-0.037432
I	= 43	0.41378

APPENDIX 2

VARIANCE-COVARIANCE MATRIX FOR PARK'S PARAMETER ESTIMATES

Copy available to DTIC does not permit fully legible reproduction

APPENDIX 3

ESTIMATED PHI MATRIX

Copy available to DTIC does not permit fully legible reproduction

ROW		COL 11	COL 12	COL 13	COL 14	COL 15	COL 16	COL 17	COL 18	COL 19	COL 20
ROW	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ROW	2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ROW	3	[illegible]	[illegible]	0.0096561	[illegible]	[illegible]	0.030125	-0.0057791	[illegible]	-0.0052355	[illegible]
ROW	4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.0053042
ROW	5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.025610	0.011268	[illegible]	0.020791	-0.0019129
ROW	6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ROW	7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ROW	8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ROW	9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.0030156
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Copy available to DTIC does not permit fully legible reproduction

		COL 31	COL 32	COL 33	COL 34	COL 35	COL 36	COL 37	COL 38	COL 39	COL 40
ROW	1	0.0031561	0.0033904	0.0029240	0.0015179	0.0007451	0.0113018	-0.0116721	0.0011661	0.0031610	0.0035509
ROW	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

		COL 41	COL 42	COL 43
ROW	1	0.0251860	-0.00014027	0.0065365
ROW	2	0.0037620	0.0001361	0.0067295
ROW	3	0.0012580	0.0001017	0.0022592
ROW	4	-0.0030187	0.0029064	0.0020053
ROW	[illegible]	[illegible]	[illegible]	[illegible]

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ONR-200-6	2. GOVT ACCESSION NO. AD-A118265	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) CONFIDENCE INTERVALS AND VALIDATION OF A FORECASTER OF QUALITY NAVY ENLISTMENTS		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Richard C. Morey and John M. McCann		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0200
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Applied Business Research, Graduate School of Business Administration, Duke University, Durham, NC 27706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 170-903, 62763N, RF 55521002
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Code 452, 800 N. Quincy Street, Arlington, VA 22217		12. REPORT DATE July, 1982
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Supported by the Naval Research Manpower R&D Program.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Confidence Intervals, Validation, Prediction, Regression, Quality Contracts

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Validation efforts and the development of rigorous statistical confidence intervals fo a non-linear predictor of quality enlistment contracts is reported. The forecaster is for Navy, male, non-prior service, active duty, High School contracts and is the one being used by the Navy Recruiting Command for use in budget determination and in goaling. The procedure deals with the complexities arising from a complicated regression model, using pooled data, i.e., colinearity, autocorrelation, heteroscedacity lagged terms, and utilizes the detailed outputs from the Park's regression package, together with Monte Carlo simulations. For 2 independent years, the model predicted within abo 3% of the actual national totals. The 90% confidence interval is ± 8% of the predicted value.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

ATE
LMED
8